Harvesting Wiki Consensus - Using Wikipedia Entries as Ontology Elements

Martin Hepp[1,2], Daniel Bachlechner[1], Katharina Siorpaes[1]

[1]Digital Enterprise Research Institute (DERI), University of Innsbruck
[2]Florida Gulf Coast University, Fort Myers, FL, USA
{martin.hepp|daniel.bachlechner|katharina.siorpaes}@deri.org

Abstract. One major obstacle towards adding machine-readable annotation to existing Web content is the lack of domain ontologies. While FOAF and Dublin Core are popular means for expressing *relationships* between Web resources and between Web resources and literal values, we widely lack unique identifiers for common concepts and instances. Also, most available ontologies have a very weak community grounding in the sense that they are designed by single individuals or small groups of individuals, while the majority of potential users is not involved in the process of proposing new ontology elements or achieving consensus. This is in sharp contrast to natural language where the evolution of the vocabulary is under the control of the user community. At the same time, we can observe that, within Wiki communities, especially Wikipedia, a large number of users is able to create comprehensive domain representations in the sense of unique, machine-feasible, identifiers and concept definitions which are sufficient for humans to grasp the intension of the concepts. The English version of Wikipedia contains now more than one million entries and thus the same amount of URIs plus a human-readable description. While this collection is on the lower end of ontology expressiveness, it is likely the largest living ontology that is available today. In this paper, we (1) show that standard Wiki technology can be easily used as an ontology development environment for named classes, reducing entry barriers for the participation of users in the creation and maintenance of lightweight ontologies, (2) prove that the URIs of Wikipedia entries are surprisingly reliable identifiers for ontology concepts, and (3) demonstrate the applicability of our approach in a use case.

Keywords. Wikis, Ontologies, Reuse, Collaborative Ontology Building, RDF, RDF-S, OWL

1. Introduction

Ontologies are consensual, explicit conceptualizations of a domain of discourse [1, 2]. In short, they are unambiguous representations of concepts, relationships between concepts (for example, but not limited to, a hierarchy), instances, and axioms. Unambiguous in this sense means two things: First, the representation should allow humans to precisely grasp the meaning of any element, so that humans have a well-

defined vocabulary at hand when annotating data, expressing queries, or drawing conclusions. Second, the representation should have a formal semantics, so that it supports machine reasoning. For a comprehensive overview, see [3]. However, it is important to note that ontologies are not just formal representations of a domain, but much more *community contracts* about such formal representations. Since a discourse is a dynamic social process, during which previous propositions are often modified, especially refined, or discarded, and new topics need to be added, such a community contract cannot be static, but must be able to reflect the community consensus at any point in time. Also, the respective community must be technically and skill-wise able to be involved in building the, or committing to the, ontology.

Ontologies can have a varying degree of expressivity, ranging from flat collections of consensual concepts to abundantly axiomatized models. Many ontologies have a subsumption hierarchy that allows to infer implicit class membership, but this is not a mandatory property. In its least expressive form, an ontology is a collection of named concepts with a natural language definition of their meaning, i.e. a controlled vocabulary.

Though more expressive ontologies support more sophisticated reasoning, even such flat ontologies can be extremely useful. Already having unique identifiers (e.g. URIs) assigned to concepts described in natural language is very beneficial, for it helps improve recall and precision in information retrieval by eliminating the significant amount that is caused by synonyms and homonyms.

Now, we can observe on one hand that there are very few real domain ontologies available; a large share of ontologies published on the Web are outdated, dead collections created in some academic research context. On the other hand, the English version of Wikipedia contains more than one million entries, which means it holds unique identifiers for the same number of concepts.

Currently, both ontology tools and ontology languages impose high entrance barriers for potential users, excluding the vast majority of Web users. The Web Ontology Language OWL [4], for example, is in several aspects non-intuitive for anybody who does not come from the Description Logics (DL) community, and publishing an ontology in a persistent manner requires infrastructure (e.g. a HTTP server) that is not available to an average user.

This in combination likely contributes to the fact that the most popular approach of creating ontologies is engineering-oriented, i.e., a small number of skilled individuals carefully constructs the representation of the domain of discourse, and releases the results at some point in time to a wider community of users. However, (1) the sequential paradigm of this approach and (2) the fact that a small group constructs the ontology for a bigger group has several weaknesses:

First, the *ontology evolution is not under the full control of the ontology user community*. For example, missing entries cannot be added by any user who reveals the need for a new concept, but has to be added by the small group of creators. This is slow and incomplete, for it may be too much a burden for the users to report missing entries. Also, the addition may take too long if the domain is undergoing conceptual change. In natural language, in comparison, the evolution of the vocabulary is under the control of the user community. Anybody can invent and define a new word or concept in the course of communications.

Second, users creating annotations cannot easily grasp the intension of a concept; there is often a lack of communication between ontology creator and user. Somebody using an ontology e.g. for annotating instances or expressing queries has little help in determining whether a given concept is suitable for his or her needs, since the formal part of the ontology only *constrains* the interpretation of a concept, but does, with the exception of very expressive ontologies, not actually *define* the meaning of this concept. This leaves the ontology user with sparse natural language descriptions, e.g. in the form of the Dublin Core field `dc:description`. Such is often not sufficient to check whether the ontology creators read the concept in the same manner as the potential ontology user does, and many ontology creators with a strong formal background put little emphasis on the natural language definitions and related non-functional properties. For example, "ice cubes" in UNSPSC can be understood as any form of ice cubes or as all ice-cube-related business documents; see [5] and [6]. As a consequence, two parties referring to the same ontology might read the intension of the concept differently, which can lead to incomplete and/or inconsistent results and operations.

We propose to directly use the infrastructure and culture of Wikis as an ontology engineering workbench that fosters true collaborative ontology creation and maintenance for lightweight ontologies, in the sense that anybody can add a new element to the ontology, and refine or modify existing ones. At the same time, we want to reuse the vast amount of Wikipedia entries (more than one million in the English version) as ontology components.

We especially propose the use of multimedia elements to improve the richness and disambiguity of informal concept definitions in an ontology. Also, we regard it as beneficial if the definition of a concept is not separated from the discussion that lead to shaping the intension of this concept, since the history of a conceptualization is a valuable part of the respective definition. In many sciences, especially philosophy, the notion of a term is hard to grasp without knowing the historical debates that lead to its introduction.

1.1 Our Contribution

In this paper, we (1) show that standard Wiki technology can be easily used as an ontology development environment without modification, reducing entry barriers for the participation of users in the creation and maintenance of lightweight ontologies, (2) present a quantitative analysis of current Wikipedia entries and their properties, (3) prove that the URIs of Wikipedia entries are surprisingly reliable identifiers for ontology concepts, and (4) demonstrate how the entries available in Wikipedia can be used as ontology elements.

1.2 Research Approach

First, we developed a minimal technical solution for using Wikipedia entries as ontology elements in RDF. Second, we took a representative, random sample (n=100) from a snapshot of the English version of Wikipedia (http://en.wikipedia.org/). Third,

we analyzed whether the concept represented by the URI at the time of adding this entry is still consistent with the most recent description retrievable at the respective URI, i.e. whether annotations made using the URI in the past remain correct despite the fact that Wiki entries can be easily modified by an open community. We especially analyzed the amount of disambiguation pages, which are inserted when the same terminology refers to distinct concepts in various contexts. Fourth, we quantitatively analyzed properties like average age of entries and amount of change per time. Since we know from statistics that random samples are, if designed properly, very reliable estimates for the full population (i.e. the full Wikipedia content), our approach returns precise data about the suitability of Wikipedia content as concepts.

1.3 Related Work

Work related to ours mainly falls into the following categories:

Community-driven Ontology Building: There is already significant literature about collaborative ontology engineering in general, e.g. Tadzebao and WebOnto (see [7]). [8] describe collaborative ontology building in analogy to Wikis, but (1) do not borrow more from the Wiki community than the pure name, (2) take a very rich ontology meta-model as the starting point, (3) do not elaborate on the community focus of ontology building, and (4) do not address the advantage of adding multimedia elements in the informal descriptions of concepts.
Wikispecies [9] can already be regarded as a first Wiki-centric ontology for species. It even includes a subsumption hierarchy; which is, however, a lesser challenge in this narrow application domain since there is a single consensual taxonomy in Biology, the Linnaean taxonomy. Recently, the term "Folksonomies" was brought up as a reference for on-the-fly classifications created by users [10, 11]. This work is very much related to ours, however there are main differences. First, we aim at reusing the vast amount of existing Wikipedia entries as ontology elements. Second, we do not distinguish between tags and Wikipedia pages, i.e. we propose to use each Wikipedia URI as the identifier for a concept. Third, we point to the importance of multimedia elements in Wikipedia entries for capturing the intension of such concepts. Fourth, we stress the fact that the history function of Wikipedia is an important component of a concept definition, since it reflects the discourse that has led to the most recent state. [12] points out that the entry barriers for ontology development and usage should be lowered. The "Simple Knowledge Organisation System (SKOS)" [13] is such an approach. [14] describes the DILIGENT knowledge processes which proposes ontology evolution and collaborative concept mapping and refinement as core techniques for building ontologies in order to deal better with domain dynamics and other ontology engineering challenges.

Augmenting Wikis with Semantic Web technology: Platypus Wiki [15] is a Wiki augmented by Semantic Web approaches, namely RDF, while we want to use Wikis for creating ontologies that can be used anywhere in the Semantic Web. [16] describes Rhizome, a Semantic Wiki system that also includes the functionality of

creating arbitrary RDF resources easily. A first version of our work has been presented in [17], but this prototype aimed at deploying a modified Wiki installation as an ontology engineering platform, while we now think that Wikipedia must be the starting point due to the enormous number of existing entries and community pickup.

At a non-ontology level, the usefulness of Wikipedia as a point of reference is discussed in [18]; however, this refers more to the aspect whether all facts said about a topic are authoritative in detail, and not whether the URIs represent consensual concepts.

Especially in the last months, there is vast interest in combining Semantic Web and Wiki approaches. However, all approaches known to us are different to ours in the way that they aim at augmenting Wikis with Semantic Web components, while we propose (1) to use unmodified Wikis as a platform for collaborative ontology building on the level of named classes, and (2) harvest the wealth of concept definitions already contained in Wikipedia. In this sense, our work is complimentary to "Semantic Web Wiki" work and can be easily combined with such approaches.

2. Understanding Wikipedia as an Ontology Asset and Ontology Workbench for the Masses

We propose to use Wiki implementations in general and especially Wikipedia as a means for

(1) defining URIs for concepts,

(2) describing the intension of those concepts in natural language, and probably augmented by multimedia elements, e.g. drawings, pictures, videos, or sound recordings, and

(3) preserving the discourse that has led to the current version of a Wiki page as an important part of the definition of the respective URI.

In a nutshell, we understand the URIs of Wiki/Wikipedia entries as identifiers for named classes. This approach appears very straightforward and might even be perceived trivial. However, it is trivial only on the technological level, but should be quantitatively validated prior to its usage.

The motivation for this approach is based on the following aspects:

(1) Wikipedia contains more than 1,000,000 entries and is likely the biggest collection of URIs augmented by a textual definition available.

(2) Wikipedia is popular as a reference and its concepts can thus be expected to have commitment by a wide audience. Based on our analysis given below, we can estimate that the total amount of active contributions (e.g. additions or modifications) exceeds 2,465,000 per month. More than 50 % of the concepts have been changed at least once per each month of their existence.

(3) Wiki technology imposes only minimal requirements on a user and is likely the simplest way of creating a persistent URI plus informal description. Anybody can add a URI for a needed concept anytime.

(4) Most Wiki packages contain a comprehensive history function that allows referring to both the latest version as well as each past version of an entry

using unique URIs. Thus, different states of the discourse become First Order Objects (FOOs) that can also be referred to.

The main paradigm of our work is simplicity, i.e. we want to support only as much functionality as can be used productively by a large share of the community.

2.1. Research Challenges

When using Wiki entries as ontology elements, we see at least the following research challenges.

Resource vs. Concept: One can argue whether the URI

```
http://en.wikipedia.org/wiki/Let_it_be
```

refers to this specific Wikipedia entry as a resource or to the respective album by the Beatles. We propose as a minimal ontological commitment to our approach that each Wikipedia entry is to be understood as the entity that an average layman associates with this description. In this sense, the URI quite naturally reflects the Beatles album, not the Wikipedia description of the album. This is a proposed social convention and can of course be debated, but makes a lot of sense in the context of our proposal.

Wiki Entries: Classes or Instances: Since there is no explicit knowledge representation model in the background, a Wiki entry can be anything; it is not clear whether it refers to an instance, a concept, or a property. By social convention, Wikipedia contains mostly entries that are proper nouns and does not include relationships and properties (see also below). So it must be clarified whether a Wiki entry is to be treated as a class or as an instance, at least if the ontology model requires a choice between these two. We solve this issue by omitting this distinction between instance and class, which is no significant problem in pure RDF or in OWL Full.

Versioning and Wiki URI Schemes: A standard Wiki already provides all functionality necessary to create a textual definition and a unique URI. For example, anybody could have added an entry for the Republic of Austria to Wikipedia, now available at

```
http://en.wikipedia.org/wiki/Austria.
```

We could immediately use this mechanism and propose to re-use this URI not only as the resource locator for retrieval of the description, but also as the identifier for the concept “Republic of Austria”. Now the problem is that since everybody can alter the text, we never know whether the current version is a monotonic extension of any previous version. So anybody who used this URI for the annotation of instances or any other statement might find that his statement no longer holds with the modified version. We propose a very hands-on solution, based on a combination of the “history” functionality in the MediaWiki distribution, and a versioning scheme embedded in the URI for concepts, same as used by the W3C for W3C documents or

the WSMO, WSMX, and WSML working groups [19]. The main idea is that the general URI, e.g.

`http://en.wikipedia.org/wiki/Austria`

always refers to the latest version, while all intermediate versions have an additional URI of their own.
In MediaWiki/Wikipedia, all intermediate versions already have unique identifiers in the following form:

`http://en.wikipedia.org/w/index.php?title=Austria&oldid=23005009`

However, since this includes the name of the script "index.php", it is not fully compliant with the design principles of URIs, see [20].
It would be desirable if the MediaWiki software is modified in a sense that makes the history entries use persistent URIs that are not bound to implementation details (e.g. PHP). This could be achieved e.g. by adding the date and time of creation (plus probably the IP address of the originator):

`http://en.wikipedia.org/wiki/Austria/YYYY-MM-DD-HH-MM-SS-IP`

This allows referring either to the latest version or to any specific version. It also makes it possible to create statements *about* a specific version. This challenge is closely related to the next one.

Conceptual Consistency of URIs over time: Wiki entries can be modified and changed by anyone and there are no substantial institutional agreements between the users who create a new entry and the ones who modify it later. It is possible that the concept represented by a URI changes substantially over time, rendering old annotations inconsistent. This is especially a problem when so called "disambiguation pages" are introduced, which happens when the community realizes that the same word is a homonym and used in very different senses in different contexts. In such cases, the original page is turned into a disambiguation page that contains separate links to the multiple context-specific entries. A core part of our work presented in this paper deals with a quantitative analysis of this problem, i.e. whether this theoretical problem is a significant obstacle, or whether it is negligible.

Dominance of Proper Nouns: While Wiki packages alone can also be used to define URIs for properties, e.g.

`http://en.wikipedia.org/wiki/isAFriendOf,`

it is by social convention that Wikipedia does not contain such entries. This means that we cannot find properties and relationships as entries in Wikipedia. There are at least three ways of dealing with this:

(1) We use properties and relationships defined in popular existing ontologies, namely Dublin Core elements [21, 22] together with Wikipedia entries.
(2) We create complimentary property ontologies in an engineering fashion, e.g. "sells", "rents", "repairs" for e-Commerce applications.
(3) We modify Wiki packages so that they can be used for defining object properties (linking resources as subjects to Wikipedia entries as objects) and

datatype properties (linking resources to literal values) and deploy a complimentary "Property Wikipedia".

All three approaches can be used in combination. We have already implemented the second and third approach. For reasons of simplicity, we restrict the example in this paper to the use of Dublin Core elements, though.

Redundancy: In collaborative ontology engineering, it can happen easily that multiple entries for the same concept are created. This has no negative impact on precision, but lowers the recall of information retrieval. Wiki contains mechanisms for merging pages in such cases. In this case, "#redirect [[PAGENAME]]" is to be inserted into the body of the discontinued page. Such links could be translated into statements of equivalence.

Selection of a Proper Ontology Meta-Model: We have to define an ontology meta-model that is suitable for a large audience. In our current approach, we support only plain RDF and completely leave out any kind of hierarchical order. This is not because we think this would be irrelevant; we are rather still researching proper support mechanisms that help yield consensual subsumption hierarchies. The problem with collaborative building of subsumption hierarchies is that a local modification can have lots of unwanted side effects that are not immediately obvious.

2.2 Example

In the following, we give an example of how Wikipedia entries can be used for describing Web resources. The example is based on the social convention that the reused Wikipedia entries are understood as the entity or concept that an average layman associates with this description, not as the Web resource itself. In this sense,

`http://en.wikipedia.org/wiki/John_Lennon`

refers to John Lennon as the singer and not to the Wikipedia entry about John Lennon.

The example below represents the facts that

- John Lennon was a contributor to the Beatles album "Let It Be",
- the title of this Beatles album is "Let It Be (Beatles Album)",
- John Lennon is related to John Lennon's discography and
- that John Lennon can be described by *"John Winston Ono Lennon was a singer, songwriter, poet and guitarist for the British rock band The Beatles"*.

```
<?xml version="1.0"?>
<!DOCTYPE rdf:RDF [<!ENTITY wiki "http://en.wikipedia.org/wiki/">]>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
 xmlns:dc="http://purl.org/dc/elements/1.1/">

<rdf:Description  rdf:about="&wiki;Let_it_be">
  <dc:title>Let It Be (Beatles Album)</dc:title>
  <dc:contributor rdf:resource="&wiki;John_Lennon"/>
</rdf:Description>

<rdf:Description rdf:about="&wiki;John_Lennon">
  <dc:description> John Winston Ono Lennon was a singer, songwriter,
poet and guitarist for the British rock band The
Beatles.</dc:description>
  <dc:relation rdf:resource="&wiki;John_Lennon_discography"/>
</rdf:Description>

</rdf:RDF>
```

Figure 1 shows the resulting RDF graph.

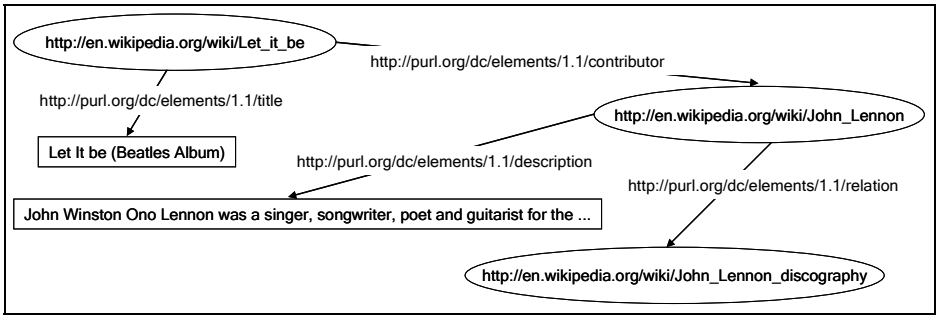

Figure 1. RDF graph of the example.

3. Evaluation

In this section, we provide evidence that our approach is not only possible from a technical standpoint, but that the URIs of Wikipedia entries are surprisingly reliable identifiers for ontology concepts, despite the fact that they are yielded in a community-driven manner.

3.1 Methodology

We want to test whether the concepts defined by the URIs of Wikipedia entries undergo significant change during their lifespan, or whether modifications tend just to add more information, which would not change the intension of the concept, but just allow additional inferences.

For this purpose, we took a random sample (n=100) of entries in the English version of WikiPedia on November 17, 2005. For this purpose, we used the „random page“

functionality of the MediaWiki software. We assume that the random number generator employed is of sufficient quality for the purpose of this survey. We know from statistics that the mean and median of a sample is a reliable estimate for the mean and median of the full population, which frees us from the need to analyse all entries in Wikipedia.

For each of the selected Wikipedia URIs, we performed the following two tasks:

(1) We compared whether the concept or entity identified by the URI has changed significantly between the very first version and the current versions, in the sense that a layman annotation of a Web resource or a layman statement about the initial concept would hold for the first version but not for the current or vice versa. We distinguished the following cases:

Case 1a: No significant change in meaning; the entry has been a stable, regular concept from its very first version to the current one.

Case 1b: The entry has always been a Wiki "disambiguation page". It refers to a stable concept (i.e. all homonyms that could be referred to by this name).

Case 2: A minor change in meaning has occurred. An example is that "Gloucester Courthouse" initially referred to the town and now refers to the "census designated place", which is still the same for many purposes.

Case 3a: There was a major change in meaning.

Case 3b: The URI was a regular entry in the beginning but turned into a disambiguation page later.

(2) For each entry, we also recorded the time and date of creation, the time and date of the last modification, the amount of editing tasks over its lifespan and per month of existence, and its age, i.e. the time lapsed between the initial creation and the date of our analysis (November 17, 2005).

Our hypothesis is that despite the ongoing change and uncontrolled editing of Wikipedia entries, there exists stable community consensus about the meaning of the respective URI.

3.3 Results

In the following, we summarize the results of our survey. Table 1 shows that only 3 % of the sample have turned into a disambiguation page during its lifespan. This is insofar important as this category of entry can have the most negative impact on precision in the usage of concepts for information retrieval, since initially, two communities might use the same URI to refer to distinct concepts.

Table 2 summarizes our findings with regard to the stability of concepts over their lifespan. One can see that 89 + 5 = 94 entries out of 100 were stable and could be used for annotation purposes without major problems. One entry underwent a slight change in meaning and 2 +3 = 5 entries were substantially modified. In other words, 95 % of the concepts can be used without or with only minor problems.

Table 1. Amount of URIs in the sample (n=100) that have turned into disambiguation pages

Disambiguation pages		
URI refers to a regular concept	**URI has always been a disambiguation page**	**URI became a disambiguation page during its lifespan**
92	5	3

Table 2. Amount of significant changes in meaning between an initial and the current version of Wikipedia entries

Significant changes in meaning between initial and current version of Wikipedia entries				
Case 1: None		Case 2: Minor	Case 3: Major	
1a: Stable, regular concept	1b: Always a disambiguation page	Slight change in meaning	3a: Major change in meaning	3b: URI became a disambiguation page
89	5	1	2	3

Table 3 shows the distribution properties of the number of modifications per Wikipedia URI. The median (i.e. the element in the middle of the sample) was changed 9.5 times during its lifespan. In other words, 50 % of the entries are changed 9.5 times or less. In relation to the duration of their existence, 50 % were changed 1.2 times a month or less. A look at the quartiles Q1 through Q4 reveals that the lowest 25 % of entries in Wikipedia was changed between 1 and 5 times (Q1), the next 25 % were changed between 5 and 9.5 times, the third 25 % were changed between 9.5[1] and 19 times and the 25 % of entries that were modified most frequently underwent between 19 and 233 modifications.

If we multiply the mean of modifications per month of existence (2.9) with the total number of Wikipedia entries at the time of the survey (850,000 in November 2005), we reveal that there are on average 2,465,000 changes to entries in the English Wikipedia each month, which points to quite an active user community.

[1] The reason why the median value is not an integer number is that we have an even sample size. In this case, if the two elements in the middle of the population have different values, per definition, the *mean* of these two is the median. Thus the 50st entry underwent 9 changes and the 51st underwent 10 changes.

Table 3. Distribution properties of the number of modifications per Wikipedia URI

Number of modifications per Wikipedia URI		
	Absolute number	Modifications per month of existence
Mean	21.8	2.9
Median	9.5	1.2
Standard Deviation	37.5	7.1
Q1	5.0	0.6
Q2	9.5	1.2
Q3	19.0	2.7
Q4	233.0	66.6

Table 4 indicates the distribution of the age of entries in days. 50 % of the entries were created less than 363 days before November 17, 2005. This is an amazing indication of how Wikipedia has gained interest and user involvement. 75 % (see the third quartile, Q3) were created less than 610 days before November 17, 2005, and only 25 % of the entries have been created more than 609 days ago.

Table 4. Lifespan in days (from creation until Nov 17, 2005)

Lifespan in days (from creation until Nov 17, 2005)	
Mean	412,6
Median	362,7
Standard Deviation	348,9
Q1	102,7
Q2	362,7
Q3	609,0
Q4	1360,7

4. Discussion

The data from our survey shows quite clearly that for the vast majority of Wikipedia entries, there is community consensus about the meaning of the URI from the very beginning to the most recent version. In other words, communities seem to be able to achieve consensus about named classes as very lightweight ontological agreements in an unsupervised fashion and with only the known mechanisms for preventing destructive changes of standard Wiki software.

As shown above, we can estimate that each month, about 2,465,000 change operations are made by Wikipedia users, but only 5 % of concepts change in a major

sense during their lifespan. We think this is a fundamental argument in favor of community-centric ontology building.
Also, our findings show that the majority of work on Wikipedia has been done in the last 20 months, since 75 % of the content of the English Wikipedia has been added in that timeframe.
Of course, there are drawbacks. First of all, what we can reuse from Wikipedia as ontology components are just named classes. There is zero support for reasoning tasks. By intention, we did not try to include any subsumption hierarchy or axioms. The reason is that other preliminary experiments which we carried out show that only very simple ontology metamodels seem suitable for collaborative ontology building. We suspect that two major reasons are the prohibitive "cost" of learning complex ontology models and the lack of transparency of effects for the average user. A simple non-consensual `rdfs:subClassOf` statement can render the annotations of a multiplicity of users incorrect, and a simple modification of `rdf:domain` can lead to class memberships that are not intended. Our future research will focus on how this skeleton can be extended towards a richer ontology meta-model without introducing new entrance barriers for users. We think for example of clever voting mechanisms with thresholds that make a subsumption relationship subject to community voting.
Also, 5% of concepts that change their meaning over time means that some annotations will become corrupt over time. However, we regard this as a trade-off decision between ontology coverage in the sense of timely addition of needed concepts, and consistency. We think that we cannot prevent the Semantic Web to break here and there. It is important to recall that a core catalyst to the success of the Web was the willingness to accept inconsistencies and broken links in return for agility and distributed evolution.
In our opinion, the delegation of ontology building to a small "elite" group of ontology engineers is conceptually flawed, since the small group has no immediate access to the representational requirements and the conceptual preferences of the community members.

5. Conclusion

We have shown that standard Wiki technology can be easily used as an ontology development environment without modification, reducing entry barriers for the participation of users in the creation and maintenance of lightweight ontologies. On the basis of a quantitative analysis of current Wikipedia entries and their properties we have provided substantial evidence that the URIs of Wikipedia entries are surprisingly reliable identifiers for ontology concepts. In addition, we have demonstrated how the more than one million entries in Wikipedia can be used as ontology elements, opening this enormous source of named classes for making the Semantic Web a reality.

Acknowledgements: The work presented in this paper is partly funded by the European Commission under the project DIP (FP6-507483), the TransIT Entwicklungs- und Transfercenter at the University of Innsbruck, and Florida Gulf Coast University.

References

[1] N. Guarino, "Formal Ontology and Information Systems," presented at FOIS '98, Trento, Italy, 1998.

[2] T. R. Gruber, "Toward principles for the design of ontologies used for knowledge sharing," *International Journal of Human-Computer Studies*, vol. 43, pp. 907-928, 1995.

[3] D. Fensel, *Ontologies: A Silver Bullet for Knowledge Management and Electronic Commerce*, 2nd ed. Berlin etc.: Springer, 2004.

[4] W3C, "OWL Web Ontology Language Guide. W3C Recommendation 10 February 2004," available at http://www.w3.org/TR/2004/REC-owl-guide-20040210/, retrieved Nov 30, 2005.

[5] M. Hepp, "A Methodology for Deriving OWL Ontologies from Products and Services Categorization Standards," presented at the 13th European Conference on Information Systems (ECIS2005), Regensburg, Germany, 2005.

[6] M. Hepp, "Representing the Hierarchy of Industrial Taxonomies in OWL: The gen/tax Approach," presented at the ISWC Workshop Semantic Web Case Studies and Best Practices for eBusiness (SWCASE05), Galway, Irland, 2005.

[7] J. Domingue, "Tadzebao and WebOnto: Discussing, Browsing, and Editing Ontologies on the Web," presented at the 11th Knowledge Acquisition for Knowledge-Based Systems Workshop, Banff, Canada, 1998.

[8] J. Bao and V. Honavar, "Collaborative Ontology Building with Wiki@nt. A multi-agent based ontology building environment," presented at the 3rd International Workshop on Evaluation of Ontology-based Tools (EON2004), Hiroshima, Japan, 2004.

[9] Wikimedia Foundation, "Wikispecies," available at http://species.wikipedia.org/, retrieved Nov 30, 2005.

[10] Wikipedia, "Folksonomy," available at http://en.wikipedia.org/wiki/Folksonomy, retrieved Nov 30, 2005.

[11] A. Mathes, "Folksonomies -Cooperative Classification and Communication Through Shared Metadata," available at http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html, retrieved Nov 30, 2005.

[12] M. D. Lytras, "Semantic web and information systems: An agenda based on discourse with community leaders," *International Journal of Semantic Web and Information Systems*, vol. 1, pp. i-xii, 2005.

[13] W3C, "Simple Knowledge Organisation System (SKOS)," available at http://www.w3.org/2004/02/skos/, retrieved Nov 30, 2005.

[14] D. Vrandecic, S. Pinto, C. Tempich, and Y. Sure, "The DILIGENT knowledge process," *Journal of Knowledge Management*, vol. 9, pp. 85-96, 2005.

[15] S. E. Campanini, P. Castagna, and R. Tazzoli, "Platypus Wiki: a Semantic Wiki Wiki Web," presented at the 1st Italian Semantic Web Workshop Semantic Web Applications and Perspectives (SWAP), Ancona, Italy, 2004.

[16] A. Souzis, "Building a Semantic Wiki," *IEEE Intelligent Systems*, vol. 20, pp. 87-91, 2005.

[17] M. Hepp, D. Bachlechner, and K. Siorpaes, "OntoWiki: Community-driven Ontology Engineering and Ontology Usage based on Wikis," presented at the 2005 International Symposium on Wikis (WikiSym 2005), San Diego, California, USA, 2005.

[18] Wikipedia, "Criticism of Wikipedia," available at http://en.wikipedia.org/wiki/Criticism_of_Wikipedia, retrieved Nov 30, 2005.

[19] J. de Bruijn and J. Kopecký, "Persistent URIs for WSMO and WSML deliverables. WSMO Note 2 February 2005," available at http://www.wsmo.org/TR/NOTE-URIs/20050202/, retrieved Nov 30, 2005.
[20] T. Berners-Lee, "Cool URIs don't change," available at http://www.w3.org/Provider/Style/URI.html, retrieved Nov 8, 2004.
[21] Dublin Core Metadata Initiative, "Dublin Core Metadata Element Set, Version 1.1: Reference Description," available at http://dublincore.org/documents/dces/, retrieved Nov 30, 2005.
[22] Dublin Core Metadata Initiative, "DCMI Metadata Terms," available at http://dublincore.org/documents/dcmi-terms/, retrieved Nov 30, 2005.